10TH INTERNATIONAL COMMAND AND CONTROL RESEARCH AND TECHNOLOGY SYMPOSIUM
The Future of C2

"The Data Warehouse in Service Oriented Architectures and Network Centric Warfare"

Topic: C4ISR Architecture
Date: June 2004

Presenter & Author: Jack Lenahan
Imagine-One Corporation
Office of The Chief Engineer
Space and NAVAL Warfare Systems Center
Charleston, South Carolina
Contact Information
Email – John.Lenahan@Navy.mil
Phone – 843-218-6080

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUN 2005		**00-00-2005 to 00-00-2005**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
The Data Warehouse in Service Oriented Architectures and Network Centric Warfare (Briefing Charts)	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Space and Naval Warfare Systems Center Charleston,PO Box 190022,North Charleston,SC,29419-9022**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**16**	

Agenda

- Introduction – Why Composeable Data Warehouse Services?
- Global and Local Data Warehouse Needs
- Definitions
- Traditional Data Warehouse
- Federated Data Warehouse
- SOA XML Web Service Warehouse Model
- CIF and GIF (Corporate Information Factories and Government Information Factories)
- Composeable Data Warehouse Services
- Results and Conclusions

Introduction – Why Composeable Data Warehouse Services?

NCW Option dominance and lockout techniques require thorough planning and adversarial understanding. The data needed to maintain superior decision processes through out a conflict is unlikely to be in one place

Process Lockout requires that a process COA output contain the necessary steps to prevent the other side from activating response processes or activating response mechanisms.

Option Dominance means that the planning and depth of understanding of the adversary's responses is so rich, that the blue plan has in advance "countered" all or most of an adversary's (the red processes) possible response options.

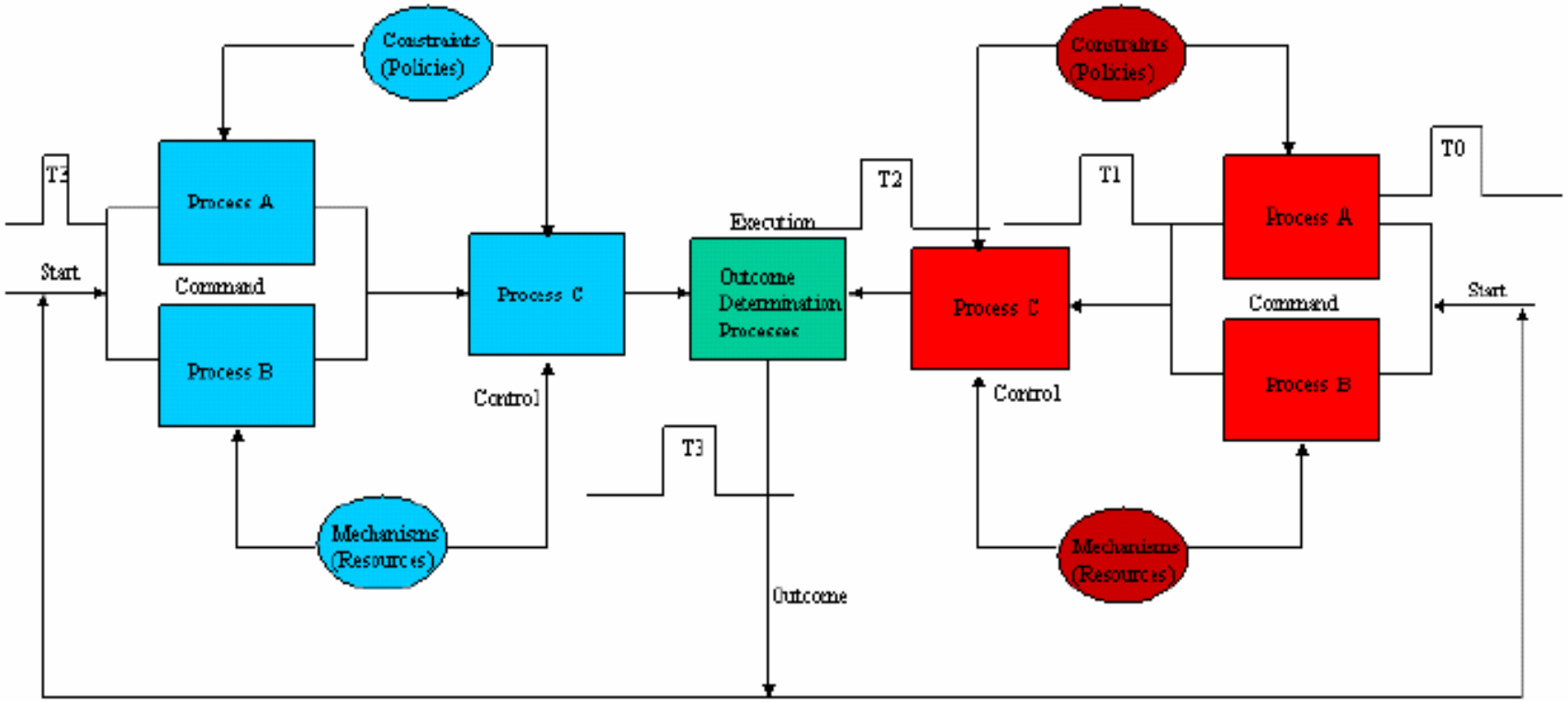
Data from multiple sources, possibly unknown prior to the start of an unexpected conflict, must be rapidly made available to decision makers in order to achieve **superior decisions**

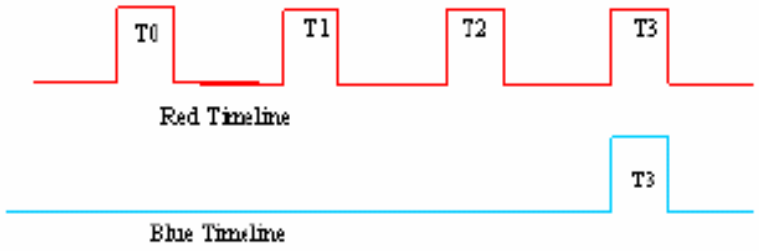
Dispersed commanders must be granted access to data sources not currently available and must be able to ask questions of that data in order to be successful

Source: " An Abstract Process and Metrics Model for Evaluating Unified Command and Control A Scenario and Technology Agnostic Approach" – Jack Lenahan
2004 Command and Control Research and Technology Symposium
"The Power of Information Age Concepts and Technologies"
June 15 – 17, 2004

Introduction Continued – Situational Awareness Model Depicting Gross lack of Data

Blue vs. Red Processes – Simple Timing Model





Time Sequence

Time 0 – Red Conceptualizes a Surprise Attack Against Blue
Time 1 – Red Completes Detailed Planning & Course of Action
Time 2 – Red Assembles Resources and Positions for Attack
Time 3 – Red Attacks- Blue not prepared Suffers Resource Damages
Score = Red +x Blue -x

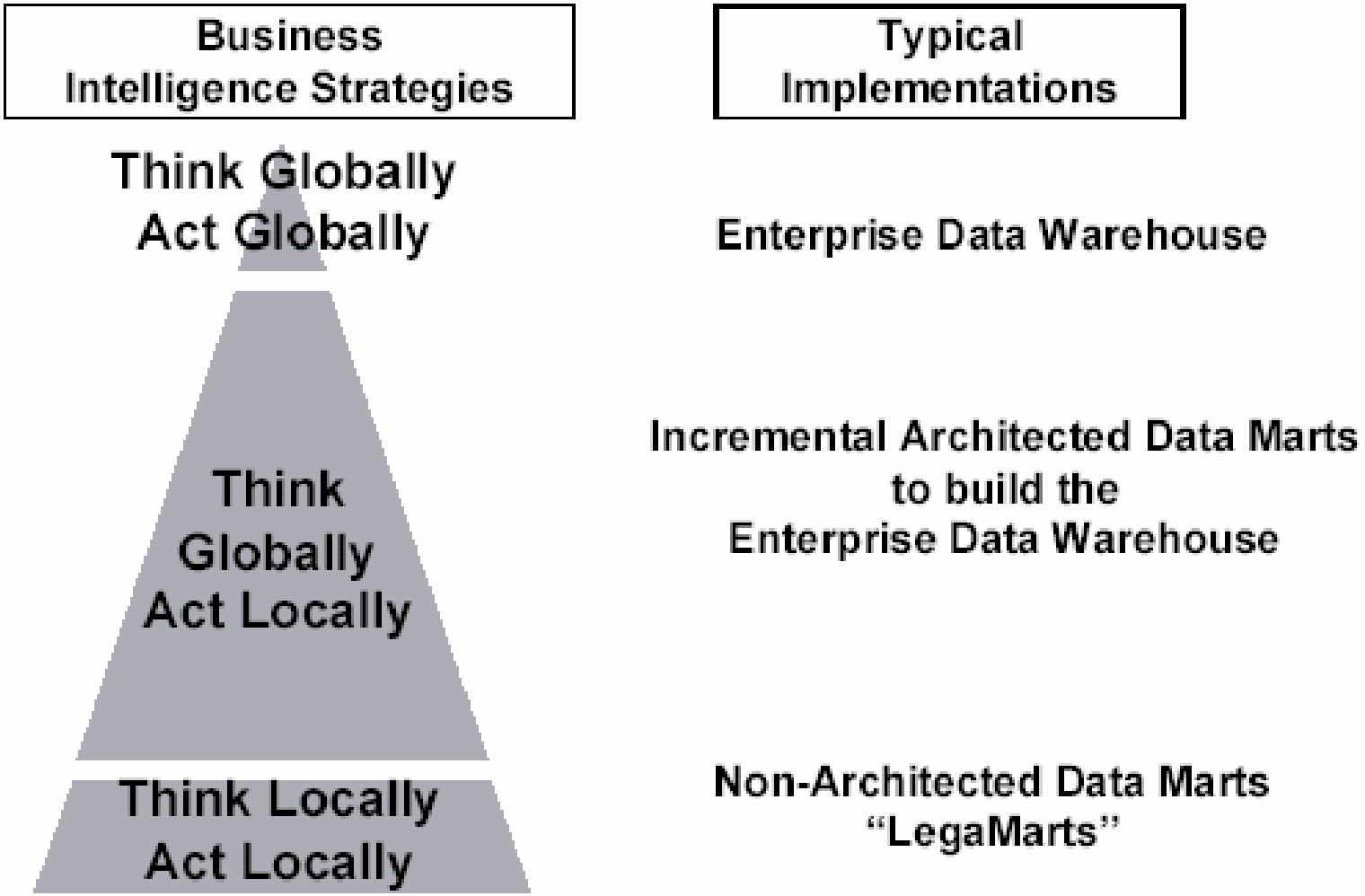
Observation of outcome

Red – Synchronized, Aware, Organized, minimum resource usage
Blue – Unorganized, Unsynchronized, Resources Lost,
Awareness begins at T3, little to no understanding

Introduction Continued – Why Composeable Data Warehouse Services?

Architects, who assume a priori, that they know what any given data user in any community of interest is going to require for day to day operations and decisions support, are arrogant and suffer from an "omniscience complex". The point of "publish all data" is to make data available to previously disenfranchised users. Composeability or the ability for an average user to define and construct (or have an intelligent assistant agent define and construct on his behalf) data warehouses, new analytics or data mining capabilities, and report types, supports the requirements of dispersed command and superior decision making at a greater level of granularity than has been available prior to this time. The capability to compose data warehouses or request that user specific analytic reports be dynamically generated offers a potentially tremendous asset to the lowest levels of command. It also permits discovery of data relationships and the uncovering of novel facts pertinent to a particular user community which would not be possible if the traditional warehouse paradigm is followed. This is true since the user community would not be at the mercy of the "omniscient architects" and thus would be free to try their own compositions. By combining GRID technology with composeable data warehouses, speed and customer interest are more likely to be satisfied.

Global Activity Planning and Analysis versus Local Activity Planning and Analysis with Legacy Data Stores (LegaMarts)



Source: "Architectures and Approaches for Successful Data Warehouses", By Douglas Hackney, April 2002, URL: http://www.egltd.com/presents/ArchitecturesApproaches.pdf

Definitions

- Citing Inmon's paper ("Inmon is the Father of the Data Warehouse"), "a (data) warehouse is a subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management's decision-making process."
  - Subject-oriented: Data that gives information about a particular subject instead of about a company's on-going operations.
  - Integrated: Data that is gathered into the data warehouse from a variety of sources and merged into a coherent whole.
  - Time-variant: All data in the data warehouse is identified with a particular time period. (Thus it cannot change after initial load and is therefore read only)

- Data warehouses are deliberately de-normalized for improving the reporting and general access speeds. Thus, they do not follow E.F. Codd's model in terms of normalization.
- A data warehouse is by definition "READ ONLY". Data mining will yield meaningless results if the underlying data content is permitted to change unexpectedly

Definitions Continued

- Extraction is defined as the process of retrieving the data to be loaded from the identified data sources in native legacy formats.
- Transformation is defined as the process of converting data from multiple sources in different formats into a single format for each field in the warehouse, prior to its being loaded.
- The operational data store is the place that extracted data is sent to and the place where the conversions are performed.
- The data warehouse is the read only relational engine where the transformed data is loaded into or stored according to the star or snowflake schemas designed to support reporting dimensions.
- Data marts are read only department or organizational level “mini-warehouses”. They are fed by the main warehouse itself. They exist to reduce the load on the central warehouse and to target data mining performance improvements due too the reduced number of rows required for departmental (not corporate level) analysis.

Definitions Continued

- **A Federated Data Warehouse Architecture is an overall system architecture that accommodates multiple DW/data mart (DM) systems, operational data stores (ODS), amorphous reporting systems, analytical applications (AAs), etc. As the Internet is a network of networks, a federated DW architecture is an architecture of architectures. It provides a framework for the integration, to the greatest extent possible, of disparate DW, DM and analytical application systems.**

Source: "Data Warehouse Delivery: Federated FAQs" - by Douglas Hackney
Published August 15, 2000
URL - http://www.datawarehouse.com/article/?articleid=2887

Corporate Information Factories and Government Information Factories Definitions and Models

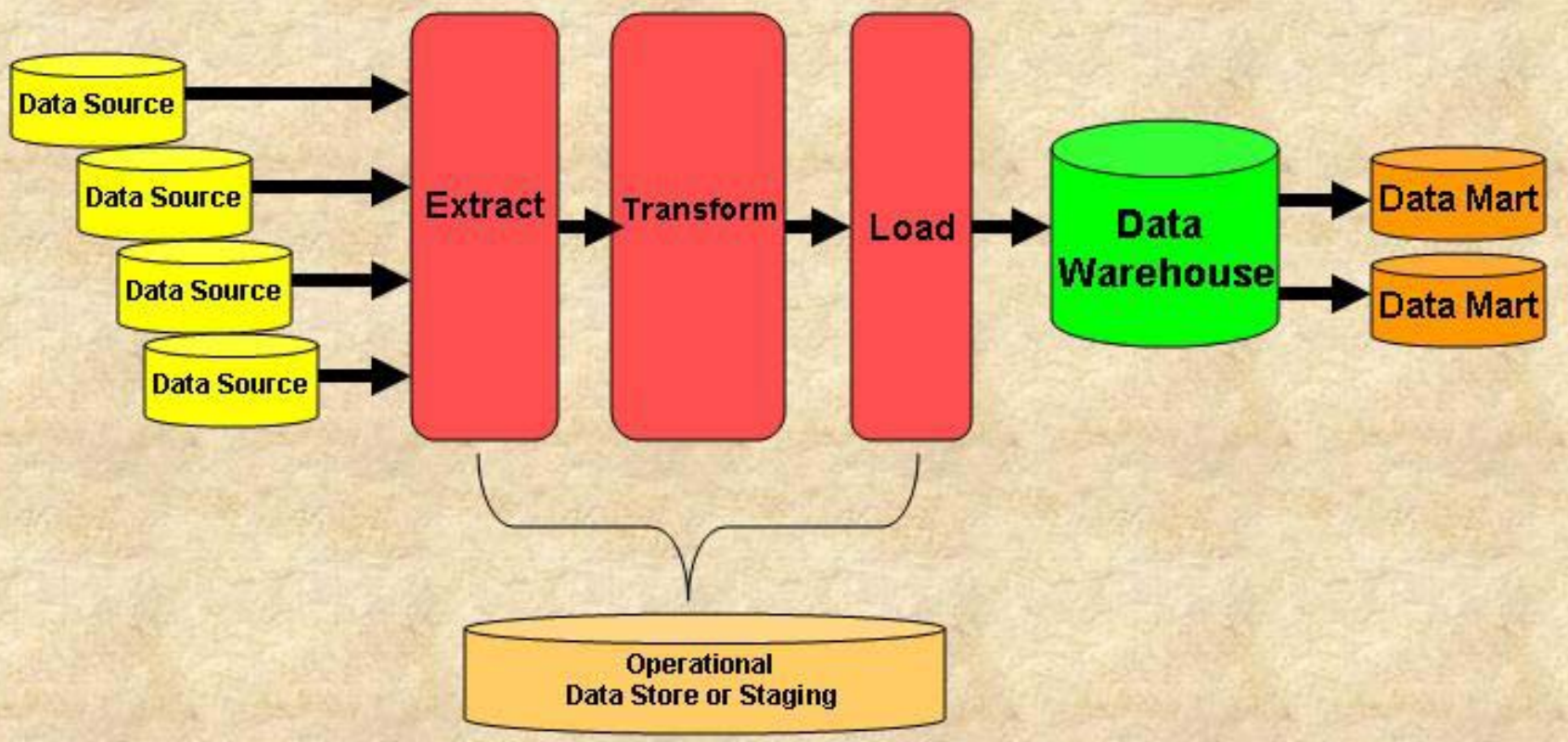
The CIF (Inmon) is an architecture for the information ecosystem, consisting of the following architectural components:

- **An applications environment**
- **An integration and transformation layer (I & T layer)**
- **A data warehouse with current and historical detailed data**
- **A data mart(s)**
- **An operational data store (ODS)**
- **An Internet and Intranet**
- **A metadata repository"**

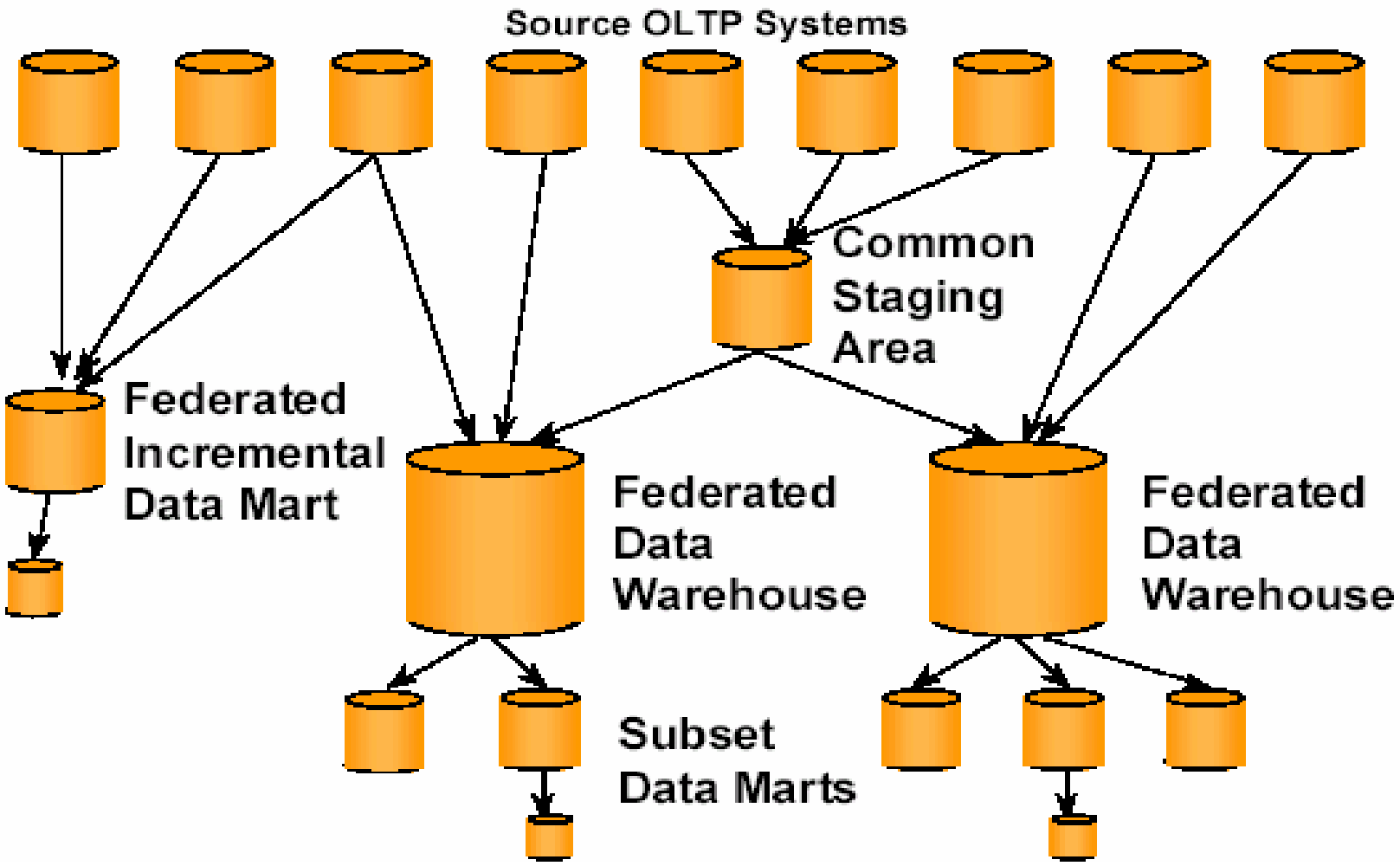
CIF/GIF Definitions Continued

- From the same source (Jens Körner) “The GIF is the counterpart of the CIF in the public sector. Inmon’s GIF is an information systems blueprint for government agencies; namely for federal, state, and local agencies, and takes into consideration the known ERP and DW needs for:
  - **Operational and informational processing**
  - **Multidimensional processing and reporting**
  - **Managing very large amounts of data**
  - **High availability**
  - **Data mining and exploration, and so forth…**
  - **In addition, the GIF takes into account the need for:**
    - **Interagency passage of data**
    - **Integrated electronic security**
    - **Predictive security (the ability to use data to anticipate threats before they occur)**
    - **Reconciliation of data**
    - **Addressing the challenges of stovepipe systems**
  - **GIF vs. CIF distinctions**
    - **(i) No concept of profits in the Public Sector**
    - **(ii) Multiple objectives in the public sector vs. single objective in the private sector**

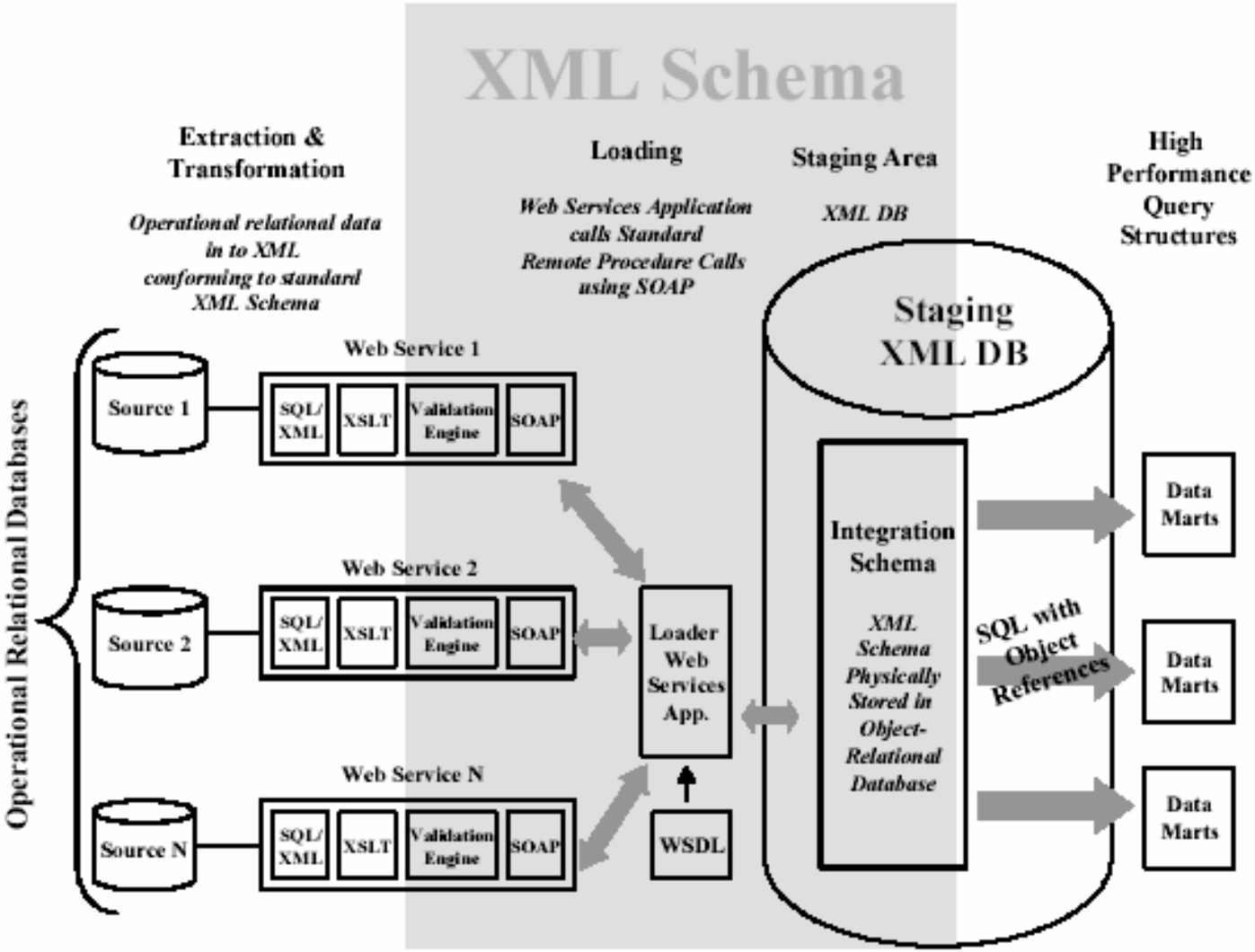
Traditional Data Warehouse Architecture



Federated Data Warehouse / Data Mart Systems

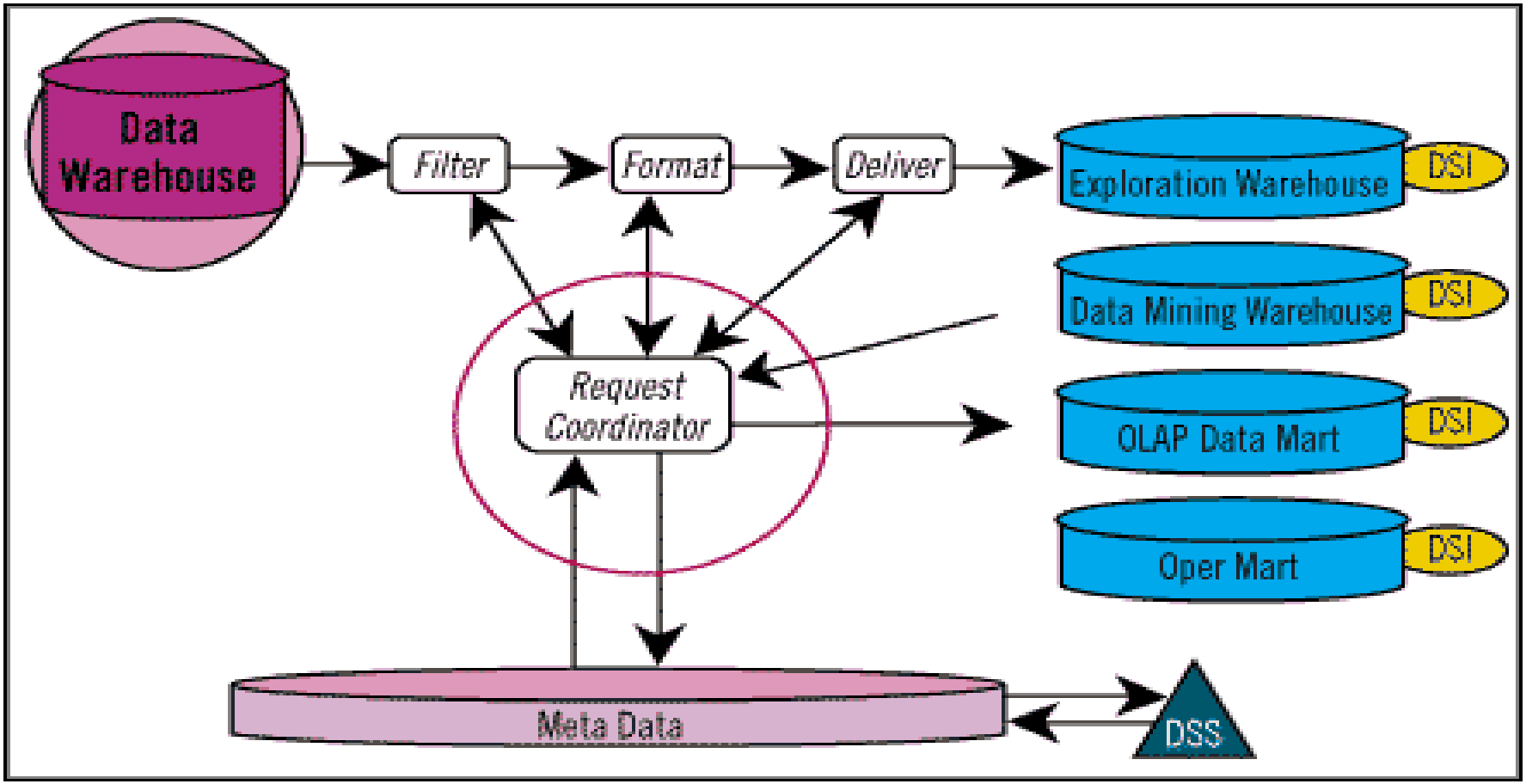


Web Services and XML Schema Based SOA Model for Agile or Composeable Data Warehouse Creation



Source: "XML Schema and Web Services for ETL in the Staging Area of a Scientific Data Warehouse" - - Mykola Dudar, Owen Eddins, et.al

Other Relevant Research – The Dynamic Creation of Marts and Reports



CIF Model with Request Manager – by Claudia Imhoff

"on demand warehouse composition services".

Results and Conclusion

The Composeable Data Warehouse offers the best methodology for achieving superior decisions at all levels of dispersed command. Composeable Data Warehouse capabilities, based upon web services, should be implemented and registered on the GIG for testing and deployment if successful. Whether the base model is an SOA XML model, a CIF, or GIF model or a combination of all these models, should be decided upon by the services or DoD as a representative of the JOINT community.